

法政大学学術機関リポジトリ
HOSEI UNIVERSITY REPOSITORY

財務会計情報の分布特性と解析手法の探求

著者	坂上 学
ページ	1-5
発行年	2018-06-20
URL	http://hdl.handle.net/10114/00022311

平成 30 年 6 月 20 日現在

機関番号：32675

研究種目：基盤研究(C) (一般)

研究期間：2013～2017

課題番号：25380620

研究課題名(和文) 財務会計情報の分布特性と解析手法の探求

研究課題名(英文) Distributon Characteristics of Financial Accounting Data and Exploring Analysis Methods

研究代表者

坂上 学 (SAKAUE, Manabu)

法政大学・経営学部・教授

研究者番号：50264792

交付決定額(研究期間全体)：(直接経費) 3,600,000円

研究成果の概要(和文)： 上場企業の資産総額、当期純利益などの財務会計数値は冪乗則に従うことを確認し、さらにこれらデータの時系列特性がフラクタル性(自己相似性)を持つことを確かめるためフラクタル次元を計算してみた。しかしサンプルサイズの問題で十分な結果が得られず、フラクタル次元については適用可能性がないことが判明した。そこで階級を色々と変化させながら頻度分析を行った結果、最頻値付近のデータ部分のみを抽出し階級値を小さくしてヒストグラムを作成して観察される分布もまた冪乗分布しており強いフラクタル性を確認することができた。この他にテキストマイニングやディープラーニングなどの新たな手法の適用可能性についても試みた。

研究成果の概要(英文)： Firstly, I confirmed the distribution characteristics of the figures such as total assets, net income, etc. of all listed companies in Japan, and also confirmed that these financial account values follow the power law. Time series characteristics of these values following power law are known to have fractal nature (self-similarity). So I have calculated fractal dimension of each financial value, but no satisfactory results are obtained due to sample size. Alternatively, I performed frequency analysis. I made histogram using part or the whole of data, then compared the shape of these histogram. Histogram using part of the data and histogram using the whole of data have very similar shape that shows both are subject to the power law. The observed distribution is power-distributed and indicated fractal nature strongly. In addition to these methods, I have also investigated the applicability of emerging methods such as Text Mining and Deep Learning.

研究分野： 会計学

キーワード： 冪分布 財務データ フラクタル性 テキストマイニング ディープラーニング

1. 研究開始当初の背景

従来の研究では、たとえば利益変動は正規分布を仮定して分析が行われてきた。しかしながら財務会計情報の主要項目についての調査によれば（坂上, 2010 & 2011）、企業規模（資産総額や売上高）や利益額について、金額と累積企業数との両対数分布を見ると、概ねマイナス1の傾きを持つ冪乗則に従う分布をしていることが明らかになっている（図1）。

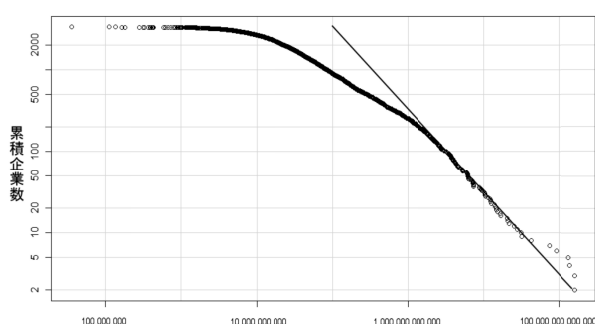


図1: 総資産額と累積企業数

企業規模が冪乗則を示すということは、古くは Simon & Ijiri (1977) でも指摘されているように特に目新しいことではないが、複数年のデータを比較しても、これらの分布の傾きが常にほぼマイナス1の傾きを示すことが注目すべき点である。統計物理学の視点から見れば、定常状態と非定常状態との相転移現象が起きていることを強く示唆しており、これらの考え方を経済現象に応用したものが経済物理学という新たな領域であるが、この経済物理学では、株価や為替の変動が正規分布ではなく、より変動幅の大きな冪分布に従い、なおかつ両対数分布ではマイナス1の傾きを示すことは、そこに安定的な状態と激しい競争状態との相転移状態が存在しているからだと主張する（高安, 2004）。別の言い方をすれば、複雑系として経済現象を捉えるという視点を導入するということである。

（引用文献）

坂上学. 2010. 「アカデミック・フォーサイト財務数値の分布特性に関する新たなパースペクティブ」『会計・監査ジャーナル』Vol. 22 No. 12, 2010 年 12 月, pp.

107-112.

坂上学 2011. 「財務数値の分布特性とベキ乗則」『会計』Vol. 180 No. 3, 2011 年 9 月, pp. 326-338.

Ijiri, Y. & Simon, H. A. 1977. Skew Distributions and Sizes of Business Firms (Studies in Mathematical and Managerial Economics Vol 24), Elsevier Science.

高安秀樹. 2004. 『経済物理学の発見』光文社.

2. 研究の目的

以上のような背景から、複雑系の現象に使われる様々な解析手法、たとえばフラクタル分析や時系列データへの主成分分析等の解析手法を経済現象に適用した場合、どのような知見が得られるであろうか。たとえば Mandelbrot & Hudson (2004) では、フラクタル幾何学の考え方を株価や為替変動の解析に適用し、「自然現象のフラクタルと経済現象のフラクタル」は同一のものであり、このフラクタル幾何学を金融市場に適用することによって、ランダムウォーク理論（変動が正規分布を示す）からは予期できない激しい動きを理解できるようになると主張している。フラクタル分析は金融データには既に適用がすすんでいるが（熊谷, 2002）これを財務会計情報に適用してみるというのが、この問題に取り組む第一歩であると考えている。

しかしながら財務会計情報の時系列データは株価や為替変動のデータに比べ、その数が極めて少なく、そのまま適用するには困難が伴う。株価や為替などでは膨大なティックデータが入手できるため、その変動のパターンを図形的に処理するフラクタル分析が可能となるが、財務会計情報で図形的なパターンを見ようとしても、圧倒的にデータ量が不足しており十分な結果が得られないという懸念がある。

そこでフラクタル幾何学をはじめとし、様々な解析手法を取り上げ、財務会計数値に当てはめてみることにし、そこから何らかの知見が得られないかを模索していくことにした。

（引用文献）

Mandelbrot, B. B. & Hudson, R. L. 2004. The Misbehavior of Markets: A Fractal View of Risk, Ruin, and Reward, Basic Books. 高安秀樹監訳. 2008. 『禁断の市場 - フラクタルでみるリスクとリ

ターン』東洋経済新報社。
熊谷善彰. 2002. 『金融時系列データのフラクタル分析』
多賀出版.

3．研究の方法

本研究を進めるにあたり、以下のような3つのステップを踏んで研究を進めた。

【ステップ1】

まず、研究をするためのインフラ作りと先行研究の文献を調査し、次年度以降の研究を進めるための基盤を整備した。

また大量の財務会計データを格納するためにデータ収集専用PCを構築し、24時間体制でEDINETより財務会計データの収集と蓄積を行った。

【ステップ2】

続いて、主として会計情報のフラクタル性についての分析を行うことにし、実際に財務会計データを用いて解析を進めることにした。まずはデータの頻度分析を行った後に、フラクタル次元の計算といったフラクタル分析の手法を用いて解析を行った。

【ステップ3】

最後に、主として当初予定していたものとは違った分析手法を試みることにした。具体的には、テキストマイニングとディープラーニングの手法を用いた分析である。

EDINETより入手できる有価証券報告書について、財務数値以外の部分についてもXBRL化が進み、手軽にテキストデータが入手できるようになったということと、ChainerやTensorFlowといったフリーのディープラーニング・フレームワークが相次いで公表され、手軽にディープラーニングの手法を応用できるための環境が整ったことによる部分が大きい。

これらの分析手法は、必ずしも冪乗分布という分布特性に対応するものではないが、新たな手法の探求ということで取り入れることにした。

4．研究成果

まず、大量の財務会計データを収集する体制を整えることにし、24時間体制でインターネットを通じ、金融庁のEDINETや東京証券取引所のTDNetより、財務会計データのみならず、さまざまな適時開示情報を収集し、蓄積を行った。EDINETのデータは日々更新されるため、データの収集にあたっては専用のコンピュータを利用して常時更新を監視しながら、リアルタイムでデータの収集を行い、分析に必要な財務会計データを、データ・サーバー上に蓄積していった。

このデータから2013年1月～2013年12月のデータを用い、当期純利益、総資本、株主資本、売上高など主要な財務数値について、その分布がどのような形状をしているかについて、調査を行い、とりわけ当期純利益の分布については、当初の予想通り典型的な冪乗分布を観察することができた。この調査で新たに発見できたことは、階級を色々と変化させながら頻度分析を行った結果、全体をカバーする極めて階級値が大きなヒストグラムを作成した場合に観察される冪乗分布が、全体のわずか2%足らずにしかない最頻値付近のデータ部分のみを抽出し、ごく小さな階級値でヒストグラムを作成した場合に観察される分布が同じく冪乗分布を示していることが判明し、強いフラクタル性（自己相似性）を確認できたことは大きな発見であった。

本研究期間中に主たる財務データの入手元であるEDINETが大幅なシステム改変を実施したことにより、従来は財務データしかXBRL化されていなかったが、有価証券報告書の全体がXBRL化され、それにより非財務情報および定性情報についてもデータが入手可能となった。そこでやや計画を変更し、非財務情報・定性情報についての分析を追加して行うこととした。

まず有価証券報告書に記載された定性的な情報（テキスト情報）の中で、「東日本大震災」に言及した企業数を調査した。対象期間は2014年1月から2014年12月までの約1年間である。その結果、「東日本大震災」という語句を2回以上

使用した企業は 236 社 (6.31%)、5 回以上使用した企業はわずかに 36 社 (0.96%) にすぎず、83.46% の企業はまったく使用していないことが判明した。これらのことから見て取れるのは、「東日本大震災」という語句の相対的な重要度の低下である。そこで TF・IDF 法によって、語句の重要度についてスコアリングして比較することにした。

調査対象期間に「東日本大震災」について言及した企業は、3,738 社中 618 社であった。したがって「東日本大震災」という語句の IDF 値は、以下の式で算出できる。

$$IDF = \log_2 \frac{3,738}{618} + 1$$

これにより IDF 値は 2.596588 と計算された。「東日本大震災」という語句が最も多く出現した企業の出現頻度 TF 値は 35 なので、当該企業における TF・IDF の値は 90.88058 と計算される。

当該企業における「原子力」や「発電」といった語句の重要度を、同じく TF・IDF 法によってスコアリングしたが、それぞれ 708.42、575.51 と算出された。つまり最も東日本大震災について言及している企業でさえ、「原子力」>「発電」>「東日本大震災」という順序となっていたことになる。

これらの結果より、上場企業全体としては、東日本大震災による直接的な損害についてはあまり影響がなかったこと、それよりも原子力発電所の事故の影響によりコスト高となった電力供給への懸念が、未だに消えていないことをうかがい知ることができたといえる。

さらに近年注目を浴びているディープラーニングの手法を適用できないか試みることにした。会計研究に対しては、従来の判別分析が用いられてきたような領域、たとえば会計不正の発見のような領域について、大量のデータを扱うことによって、より精度の高い判定を行える可能性があることが分かった。

しかしながらディープラーニングを実行するためのハードウェア的な制約 (Cuda コアを大量

に持つプロセッサが品薄で入手が大幅に遅れたこと) や Chainer や TensorFlow に代表されるディープラーニング・フレームワークのインストールと環境整備に時間がかかったため、研究期間が終わるまでに実際の検証作業を実施することができなかった。

この他の分析手法としては、いわゆる p 値による仮説検定手法についても、近年問題視されることが多く、ベイズ推定や生物学的同等性の検証方法の有用性についても研究を行った。

5 . 主な発表論文等

〔雑誌論文〕(計 6 件)

1. 坂上学、2017、「会計研究におけるディープラーニングの適用可能性」『産業経理』第 77 巻第 3 号、84-97 頁、査読なし。
2. 坂上学、2016、「XBRL 情報の投資情報としての高度化 次世代 EDINET における InlineXBRL とディメンション技術の概要」『証券アナリストジャーナル』第 54 巻第 1 号、18-27 頁、査読なし。
3. 坂上学、2016、「過渡期にある会計研究」『會計』第 189 巻第 4 号、395-405 頁、査読なし。
4. 坂上学、2015、「監査データ標準 (ADS) をめぐる動向について - XBRL GL 技術の監査への応用 - 」『会計監査ジャーナル』第 716 号、117-124 頁、査読なし。
5. Zhenyang, B. Sakaue, M. & Takeda, F. 2014. “The Impact of XBRL Adoption on the Information Environment: Evidence from Japan,” *The Japanese Accounting Review*, Vol. 4, pp. 49-74、査読あり。
6. 坂上学、2014、「冪乗則と非線形モデルへの展開」『會計』第 185 巻第 3 号、15-27 頁、査読なし。

〔学会発表〕(計 6 件)

1. 坂上学、「帳簿の電子化と複式簿記の役割」日本簿記学会第 33 回関西西部会統一論題『電

- 子社会の本格化と複式簿記の終焉』、2017 年 5 月 20 日、中部大学 (愛知県春日井市)。
2. 坂上学、「未来のディスクロージャーとその分析」日本経営分析学会第 34 回全国大会統一論題『パブリック・ディスクロージャーと経営分析』第 4 報告、2017 年 4 月 30 日、関西大学 (大阪府吹田市)。
 3. 坂上学、「事象アプローチと資金会計論との接点」、資金会計フォーラム統一論題報告、2016 年 11 月 19 日、愛知学院大学 (愛知県名古屋市)。
 4. 坂上学、「過渡期にある会計研究」解題、日本会計研究学会関東部会統一論題、2015 年 12 月 5 日、東京経済大学 (東京都国分寺市)。
 5. 坂上学、「テキストマイニングとディスクロージャー研究」日本ディスクロージャー研究学会第 10 回研究大会統一論題報告『ディスクロージャー研究の新動向』第 1 報告、2014 年 12 月 20 日、名古屋市立大学 (愛知県名古屋市)。
 6. 坂上学、「冪乗則と非線形モデルへの展開」日本会計研究学会第 72 回全国大会統一論題報告『会計学のイノベーション』第 3 報告、2013 年 9 月 5 日～2013 年 9 月 6 日、中部大学 (愛知県名古屋市)。

〔図書〕(計 1 件)

坂上学、2016、『事象アプローチによる会計ディスクロージャーの拡張』、総頁数 283 頁、中央経済社。

〔その他〕

ホームページ等：

<http://www.sakaue-lab.org/>

6 . 研究組織

(1) 研究代表者

坂上 学 (SAKAUE, Manabu)

法政大学・経営学部・教授